

Modelos de Supervivencia

Daniel Villa

Tabla de contenidos

Base de datos inicial:	1
Modelo Kaplan-Meier:	2
CCM & TREAT	2
Modelo COX	4
Conclusiones	8

Base de datos inicial:

Se muestra una base de datos inicial para ejemplificar como quedo esta para el trabajo con el respectivo modelo.

```
data.model %>% head() %>% kable()
```

	treat	totno	tyoj	tld	restt	sense	marga	anatf	time_anatf	ccm	time_ccm
P-1	1	3	1	1	1	1	1	0	24	0	24
P-2	1	2	1	2	2	1	1	0	24	0	24
P-3	1	4	1	1	3	1	1	0	24	0	24
P-4	1	1	1	1	1	1	1	0	24	0	24
P-5	2	2	1	1	2	1	1	0	24	0	24
P-6	2	7	2	2	6	1	1	0	24	0	24

```
# Tomamos las variables que cambian a través del tiempo y su respectiva supervivencia  
df <- data.model[,c(1,8,9,10,11)]  
df %>% head() %>% kable()
```

	treat	anatf	time_anatf	ccm	time_ccm
P-1	1	0	24	0	24
P-2	1	0	24	0	24
P-3	1	0	24	0	24
P-4	1	0	24	0	24
P-5	2	0	24	0	24
P-6	2	0	24	0	24

Después de haber seleccionado las variables que nos interesan pasamos a la creación y ajuste del modelo.

Vemos que las variables a que nos quedaron fueron:

```
names(df)
```

```
[1] "treat"      "anatf"      "time_anatf" "ccm"        "time_ccm"
```

Modelo Kaplan-Meier:

CCM & TREAT

Creemos un modelo kaplan-meier y veremos su resultado:

Este modelo se ajusto con las variables de CCM (Presencia de Caries) & TREAT (tipo de tratamiento)

```
# Ajuste del modelo Kaplan:
fit <- survfit(Surv(time_ccm, ccm) ~ treat, data = df)

summary(fit)
```

```
Call: survfit(formula = Surv(time_ccm, ccm) ~ treat, data = df)
```

```

          treat=1
      time      n.risk      n.event      survival      std.err lower 95% CI
24.0000      98.0000       2.0000       0.9796       0.0143      0.9520
upper 95% CI
1.0000

          treat=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

12	111	1	0.991	0.00897	0.974	1.000
18	110	1	0.982	0.01263	0.958	1.000
24	109	2	0.964	0.01769	0.930	0.999

Vemos aquí un resumen del modelo que es de dos curvas con sus respectivos eventos...

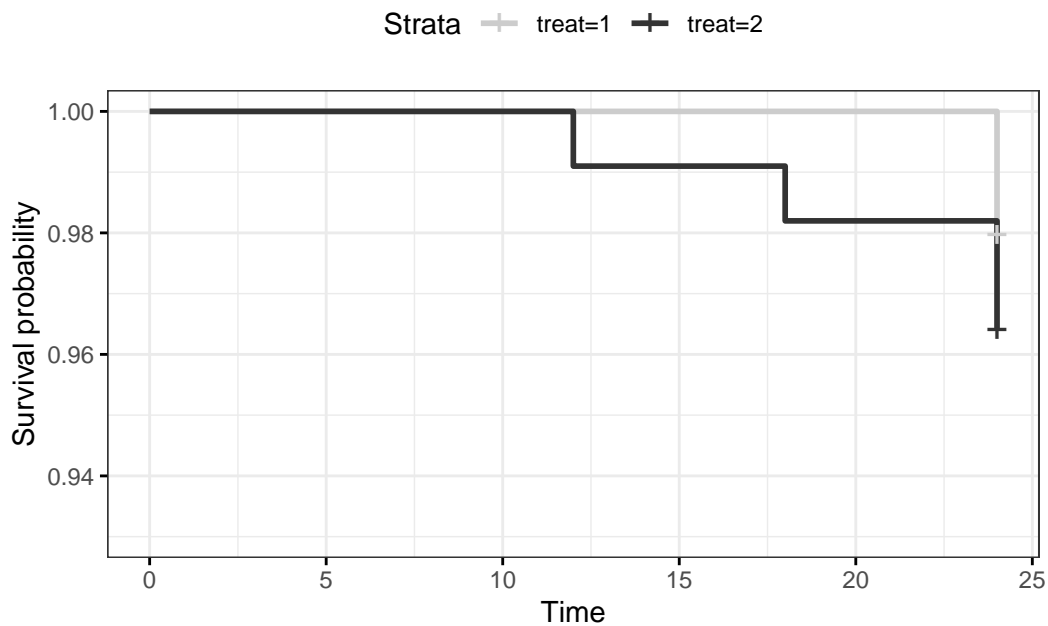
Dado la falta de eventos, los datos se toman como censurados hacia la derecha por presentar la posibilidad de fallar después del tiempo del experimento, por lo que más adelante ajustaremos un modelo COX para ver si la presencia de otras variables afecta significativamente los resultados.

Siguiendo con el resumen, tenemos que para la curva del **ART** solo presenta dos eventos por lo cual la tasa de supervivencia es muy alta, tanto que su su intervalo con una confianza del 95% en su limite superior alcanza valores muy cercanos a 1, es decir, al cabo de 2 años los usuarios del *ART* NO presentarían caries en condiciones iguales a la base del modelo y para el caso de la **AMALGAMA** es “casi” igual; se presentan más casos de caries en diferentes periodos de tiempo desde los 12 meses a los 24 meses, esto solo nos puede decir que la tasa de supervivencia también es alta por la poca cantidad de datos de pacientes que pasaron de un estado sano a presentar caries en los dientes intervenidos, por ende solo podemos decir:

Las curvas de supervivencia de los dos tratamientos son casi lineales (esto se debe por la efectividad de los tratamientos) pero realmente si se busca uno en las cuales se le pueda asegurar a los pacientes u en la toma de decisión a la hora de escoger entre estos dos, se tomaría el ART dado que los casos de falla representan el 0.02% de la población a los cuales se les aplico este tratamiento versus la AMALGAMA que el 0.04% de la población tratada, es decir, el doble de casos, para simplificarlo: “por cada 10 pacientes tendríamos al menos un paciente con caries” (ART) - “por cada 10 pacientes tendríamos al menos dos pacientes con caries” (AMALGAMA)

comprobamos esto con su respectiva curva:

```
ggsurvplot(fit,data = df,
            ylim = c(0.93,1),palette = "grey", ggtheme = theme_bw())
```



Aquí corroboramos lo dicho anteriormente, por lo que en el caso del ART solo se presentó una falla en el mes #24 (podríamos pensar que sería un error humano), es decir que el paciente por XY variables exógenas al estudio afectó su diente y así decir que este paciente es un “outliers”.

Otro efecto que no se mide es la mediana del modelo debido a que la “curva” no cumple con las condiciones por ende los valores aproximados son ∞ (una razón más para tener un modelo con covariables).

Nota: se omitió el modelo con la variable `anatf` por presentar valores similares a los `ccm`

Modelo COX

Ahora creamos el modelo con covariables para ver si estas explican mejor la supervivencia frente al *ART VS AMALGAM*

```
# Ejemplo Primario
fit2 <- coxph(Surv(time_ccm,ccm)~treat+totno+tyoj+tld+restt+sense,data=data.model2)

fit2 %>% summary()
```

	time	n.risk	n.event	n.censor	surv	std.err	upper	lower	strata
1	24.00	98.00	2.00	96.00	0.98	0.01	1.00	0.95	ART
2	12.00	111.00	1.00	0.00	0.99	0.01	1.00	0.97	AMALGAM
3	18.00	110.00	1.00	0.00	0.98	0.01	1.00	0.96	AMALGAM
4	24.00	109.00	2.00	107.00	0.96	0.02	1.00	0.93	AMALGAM

Vemos que son pocas las variables que son significativas en comparación a su categoría de referencia, por eso y tomando el test de verosimilitud que nos dirá si este modelo según el test global que no es útil para explicar la supervivencia de los sujetos.

Ahora utilizaremos el método de Stepwise para ver si hay algún modelo que nos diga o explique mejor que el anterior:

```
formula = Surv(time_ccm,ccm)~treat+totno+tyoj+tld+restt+sense+marga

paso.cox <- stepwiseCox(formula,
  na.omit(data.model2), include=NULL,
  selection=c("bidirection"), select="AIC",
  method=c("efron"), weights = NULL, best = NULL)

# Mejor modelo
paso.cox$Coefficients
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
marga2	-15.779994	1.402283e-07	16766.181160	-0.00094118	9.992490e-01
marga3	5.937212	3.788773e+02	1.126857	5.26882419	1.373004e-07

Al ver que nuestro mejor modelo para explicar las caries es con las covariables TREAT & MARGA, entonces vemos que:

```
fit2 <- coxph(Surv(time_ccm,ccm)~treat+marga,
  data=na.omit(data.model2))

fit2
```

Call:

```
coxph(formula = Surv(time_ccm, ccm) ~ treat + marga, data = na.omit(data.model2))
```

	coef	exp(coef)	se(coef)	z	p
treatAMALGAM	5.605e-01	1.752e+00	8.686e-01	0.645	0.519

```
marga2      -1.593e+01  1.210e-07  1.683e+04 -0.001    0.999
marga3       5.938e+00  3.793e+02  1.128e+00  5.263  1.41e-07
```

```
Likelihood ratio test=37.41 on 3 df, p=3.761e-08
n= 208, number of events= 6
```

El hazard rate asociado con el factor (columna `exp(coef)`) es mayor que 1 (en el caso de `treat`) lo que indica que el ART tienen más supervivencia que los pacientes con AMALGAM; para el caso de `marga2` este representa una mayor supervivencia con el dato del factor siguiente, es decir, el valor #3.

Ahora los valores P muestran la poca significancia del modelo, pero esto no es relevante dado a que necesitamos evaluar si el test de bondad y ajuste del modelo (“likelihood ratio test”) resulta significativo indicando que el tratamiento y la adaptación marginal contribuyen claramente a explicar la supervivencia de los sujetos del estudio.

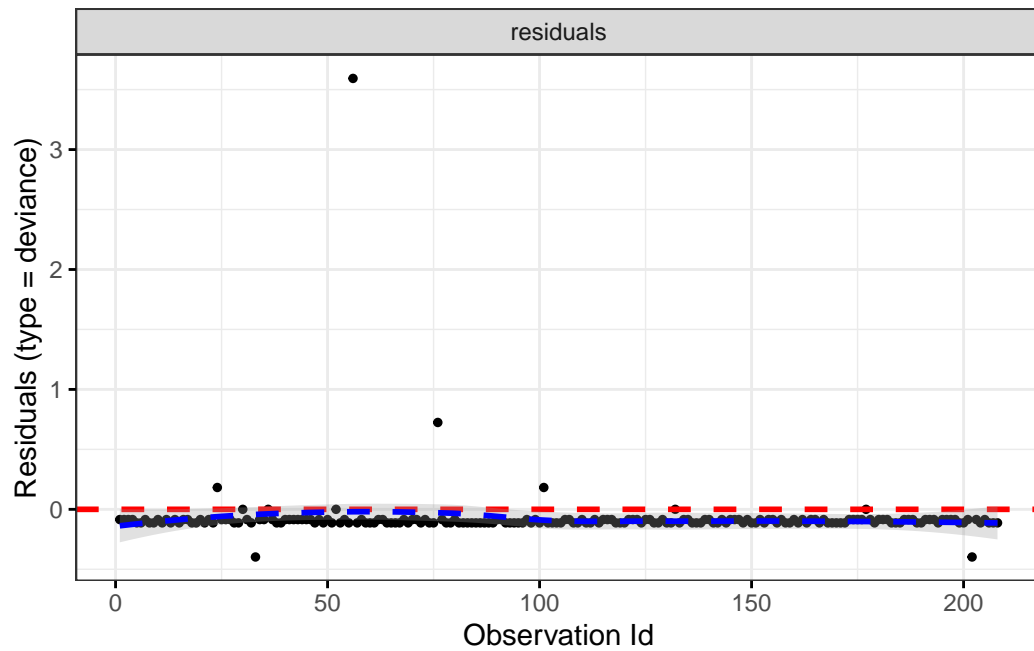
Probamos supuestos del modelo:

```
cox.zph(fit2)
```

```
      chisq df    p
treat  0.974  1 0.324
marga  8.893  2 0.012
GLOBAL 10.042  3 0.018
```

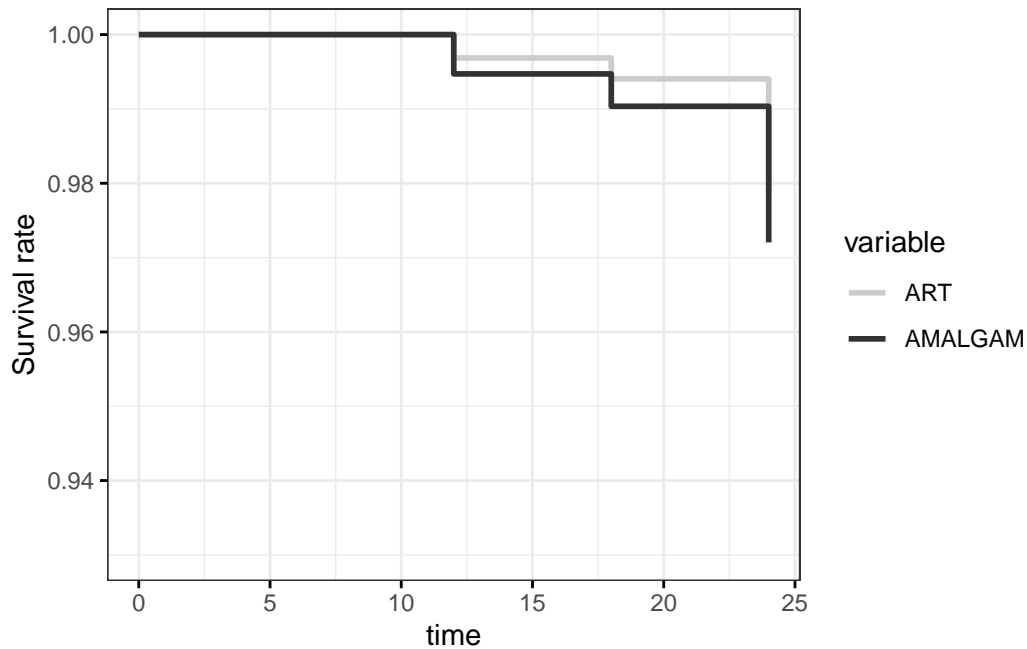
Dado que el pvalor asociado del test resulta significativo podemos descartar a modo global la hipótesis de riesgos proporcionales asociados con la variable `treat`.

```
ggcoxdiagnostics(fit2, type = "deviance",
                 linear.predictions = FALSE)
```



Como podemos ver hay datos salidos tales como el 52 o 56, pero estos datos no se pueden imputar por la relevancia del cambio en los datos para el modelo.

```
ggadjustedcurves(fit2, data=na.omit(data.model2),
  variable = "treat",
  palette = "grey", ylim = c(0.93,1), ggtheme = theme_bw())
```



Observando la curva podemos ver como afecta un poco la supervivencia más que todo del ART al punto de tener valores cruzados con los del AMALGAMA.

Conclusiones

El modelo COX no cumplió con lo esperado, pero no fue malo del todo, llevando a una conclusión algo obvia y es que hay mayor supervivencia si la adaptación marginal esta bien hecha, por lo cual se descarta la variable y volvemos al modelo inicial el cual nos arroja que la supervivencia en términos de efectividad se la lleva casi por el doble el ART teniendo tasas, pero esto depende a su vez de los cuidados futuros y el desarrollo clínico en la cita del paciente.

El modelo Kaplan es mucho mejor a pesar de no tener el registro de la mediana.